

# Bayesian model comparison in cosmology

Daniel J. Mortlock

Astrophysics Group, Blackett Laboratory, and Department of Mathematics  
 Imperial College London, London SW7 2AZ, United Kingdom  
 email: [mortlock@ic.ac.uk](mailto:mortlock@ic.ac.uk)

**Abstract.** The standard Bayesian model formalism comparison cannot be applied to most cosmological models as they lack well-motivated parameter priors. However, if the data-set being used is separable then it is possible to use some of the data to obtain the necessary parameter distributions, the rest of the data being retained for model comparison. While such methods are not fully prescriptive, they provide a route to applying Bayesian model comparison in cosmological situations where it could not otherwise be used.

**Keywords.** Bayesian inference, cosmology, statistical techniques

## 1. Introduction

Much of observational cosmology can be thought of as an attempt to use astronomical data to discriminate between the different cosmological models under consideration. Given both the inevitably imperfect data and the intrinsically stochastic nature of many cosmological measurements (*i.e.*, cosmic variance), it is generally impossible to come to absolute conclusions about the various candidate models; the best that can be hoped for is to evaluate the probabilities, conditional on the the available data, that each of the candidate models is the correct description of the Universe. The fact that there is, as far as is known, just a single observable Universe (*i.e.*, there is no ensemble from which it has been drawn), means that such probabilities cannot be frequency-based, and must instead must represent a degree of implication. Self-consistency arguments then require (Cox 1946) that these probabilities be manipulated and inverted using Bayes's theorem.

Taken together, the above facts imply that Bayesian model comparison (Section 2) should be used to assess how well different cosmological models explain the available data, although the fact that most such models have unspecified parameters is a significant difficulty for this approach (Section 3). This problem can be solved for separable data-sets as it is possible to use a two-step method of model comparison (Section 4), illustrated here with high-redshift supernova (SN) data (Section 5).

## 2. Bayesian model comparison

Given that one of a set of  $N$  models,  $\{M_1, M_2, \dots, M_N\}$ , is assumed to be true, the state of knowledge conditional on all the available (and relevant) information,  $I$ , is fully summarised by the probabilities  $\Pr(M_1|I), \Pr(M_2|I), \dots, \Pr(M_N|I)$ , where  $\Pr(M_i|I)$  is the probability that the  $i$ 'th model is correct (and  $i \in \{1, 2, \dots, N\}$ ). In the light of some new data,  $d$ , that has not already been included in the above probabilities, Bayes's theorem gives the updated probability that model  $i$  is correct as

$$\Pr(M_i|d, I) = \frac{\Pr(M_i|I) \Pr(d|M_i, I)}{\sum_{j=1}^N \Pr(M_j|I) \Pr(d|M_j, I)}, \quad (2.1)$$

where  $\Pr(d|M_i, I)$  is the marginal likelihood under model  $M_i$ .

If model  $M_i$  has  $N_i$  unspecified parameters  $\{\theta_i\} = \{\theta_{i,1}, \theta_{i,2}, \dots, \theta_{i,N_i}\}$  then the model-averaged likelihood is obtained by marginalising over these parameters to give

$$\Pr(d|M_i, I) = \int \Pr(\{\theta_i\}|M_i, I) \Pr(d|\{\theta_i\}, M_i, I) d\theta_{i,1} d\theta_{i,2} \dots d\theta_{i,N_i}, \quad (2.2)$$

where  $\Pr(\{\theta_i\}|M_i, I)$  is the prior distribution of the parameter values in this model. This expression demonstrates that the full specification of a model requires not just an explicit parameterisation, but a distribution for those parameters as well; two mathematically identical descriptions with different parameter priors are, in fact, different models.

### 3. Comparison of models without parameter priors

Equations 2.1 and 2.2 together summarise a self-consistent method for assessing which of a set of models is better supported by the available information, provided that the parameter priors for all the models are explicitly defined and unit-normalised. In particular, while it is often possible to obtain sensible parameter constraints based on an improper prior, such as  $\Pr(\{\theta_i\}|M_i, I)$  constant for all  $\{\theta_i\}$ , the resultant marginal likelihood is meaningless (Dickey 1961). Unfortunately, it is commonly the case in astronomy and cosmology that there is no compelling form for the models' parameter priors and, further, that the natural uninformative prior distributions are improper and cannot be normalised. The apparent implication is that Bayesian model comparison, at least in the form described in Section 2, cannot be used in cosmology, an idea that has been explored previously by, *e.g.*, Efstathiou (2008) and Jenkins & Peacock (2011). The disturbing corollary would be that there is no self-consistent method to choose between the available cosmological models, even if they are completely quantitative and mathematically well-defined.

### 4. Model comparison with separable data

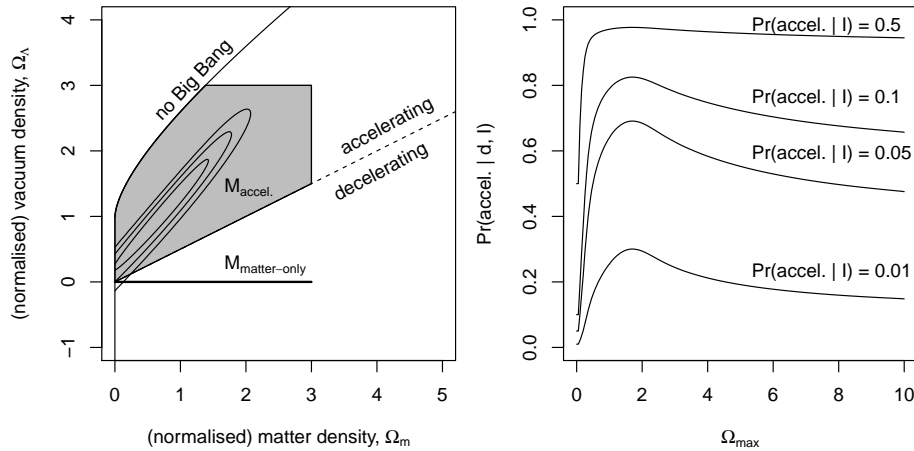
The idea that the relative degree of support for models with unspecified parameters is undefined is at odds with the marked – and data-driven – progress that has been made in cosmology over the last century. Clearly it *is* possible to use data to choose sensibly between models even if they do not have well-motivated parameter priors; but can this be formalised in a way that satisfies Bayes's theorem and is hence logically self-consistent?

One possibility is, for separable data-sets (such as those which consist of measurements of many astronomical sources), to use some of the available data to obtain the necessary parameter priors and to then use the remaining data for model comparison. This is an old concept, dating back at least to Lempers (1971) and explored subsequently by, *e.g.*, Spiegelhalter & Smith (1982) and O'Hagan (1995). The central idea is to partition the data as  $d = (d_1, d_2)$ , with the first partition of training data used to obtain the (partial) posterior distribution for the parameters of  $i$ 'th model as

$$\Pr(\{\theta_i\}|d_1, M_i, I) = \frac{\Pr(\{\theta_i\}|M_i, I) \Pr(d_1|\{\theta_i\}, M_i, I)}{\int \Pr(\{\theta'_i\}|M_i, I) \Pr(d_1|\{\theta'_i\}, M_i, I) d\theta'_{i,1} d\theta'_{i,2} \dots d\theta'_{i,N_i}}, \quad (4.1)$$

where  $\Pr(\{\theta_i\}|M_i, I)$ , which need *not* be normalisable, should be a highly uninformative prior. This posterior distribution can then be used as the prior needed to obtain a meaningful marginal likelihood, which can then be evaluated for the testing data as

$$\Pr(d_2|d_1, M_i, I) = \int \Pr(\{\theta_i\}|d_1, M_i, I) \Pr(d_2|\{\theta_i\}, M_i, I) d\theta_{i,1} d\theta_{i,2} \dots d\theta_{i,N_i}. \quad (4.2)$$



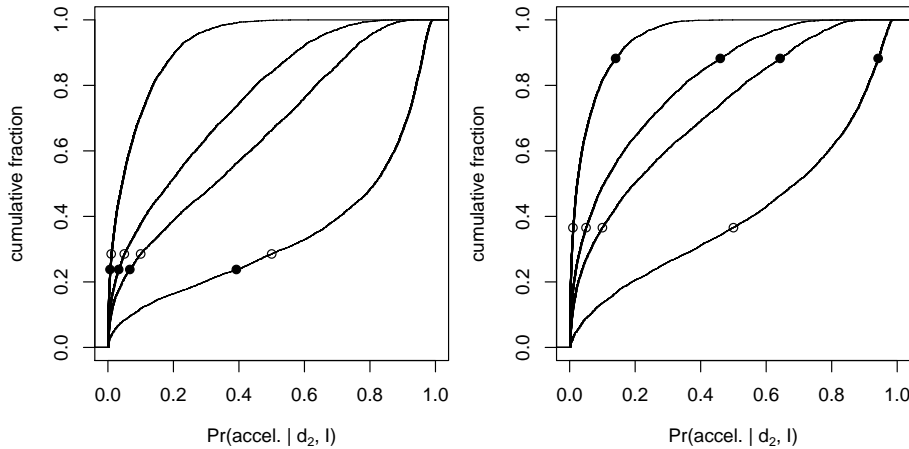
**Figure 1.** (left) The posterior distribution of  $\Omega_m$  and  $\Omega_\Lambda$  implied by the Perlmutter *et al.* (1999) SCP SN data and a uniform prior with  $\Omega_m \geq 0$ . Highest posterior density contours enclosing 68.3%, 95.4% and 99.7% of the posterior probability are shown. Also shown are the prior distributions of the accelerating model and matter only model for  $\Omega_{\max} = 3$ . (right) The dependence of  $\Pr(\text{accel.} | d, I)$  on  $\Omega_{\max}$ , shown for different prior probabilities,  $\Pr(\text{accel.} | I)$ .

This marginal likelihood is coherent, in the sense that it provides self-consistent updated posterior probabilities when inserted into Equation 2.1, but there is also ambiguity in how to partition the data: there is no compelling scheme for partitioning the data. It is tempting to average over the possible partitions, but this approach does not have a rigorous motivation. Despite these ambiguities, this two-step method of Bayesian model comparison for separable data does satisfy the Cox (1946) self-consistency requirements and so provide a means of calculating posterior probabilities for cosmological models with unspecified parameter priors.

## 5. Example: late-time acceleration and supernovae

One of the most significant recent cosmological discoveries was that the Universe's expansion rate is increasing, a result which is often linked most strongly to the observations of distant SNe made by Riess *et al.* (1998) and Perlmutter *et al.* (1999). The comparative faintness of the SNe, given their redshifts and light-curve decay timescales, indicated that the (normalised) cosmological constant,  $\Omega_\Lambda$ , is sufficiently large to override the deceleration caused by the (normalised) matter density,  $\Omega_m$ . Riess *et al.* (1998) and Perlmutter *et al.* (1999) used their SNe measurements,  $d$ , to obtain posterior distributions of the form  $\Pr(\Omega_\Lambda, \Omega_m | d, I)$ , under the assumption of uninformative (and improper) uniform priors of the form  $\Pr(\Omega_m, \Omega_\Lambda) \propto \Theta(\Omega_m)$ , where  $\Theta(x)$  is the Heaviside step function. The posterior distribution for the 42 SCP SNe from Perlmutter *et al.* (1999), reproduced in Figure 1, reveals that most of the models that are consistent with the data correspond to an accelerating universe (*i.e.*,  $\Omega_\Lambda > \Omega_m/2$ ).

But do these data provide *quantitative* evidence of cosmological acceleration? Riess *et al.* (1998) approached this question by calculating the fraction of the posterior with  $\Omega_\Lambda > \Omega_m/2$ , which is an apparently compelling 0.997 for the case shown in Figure 1. The relevant Bayesian calculation (*c.f.* Drell *et al.* 2000) should, however, be based on the marginal likelihoods of an accelerating model (for which the prior is non-zero only for  $\Omega_\Lambda > \Omega_m/2$ ) and a decelerating model (for which the obvious option is a matter-only model with  $\Omega_\Lambda = 0$ ). Such models can be fully specified (in the sense defined in Section 2)



**Figure 2.** The distribution of  $\text{Pr}(\text{accel.}|d_2, I)$  obtained from different partitions of the Perlmutter *et al.* (1999) SN data set with training sets of 10 (left) and 21 (right) SNe. The open symbols indicate the prior values (of, from left to right, 0.01, 0.05, 0.1 and 0.5) and the solid symbols show the posterior values given by training and testing samples that alternate in redshift.

by adding the restrictions that  $0 \leq \Omega_m \leq \Omega_{\text{max}}$  and  $0 \leq \Omega_\Lambda \leq \min[\Omega_{\text{max}}, \Omega_{\Lambda, \text{BB}}(\Omega_m)]$  (defined to reject models that did not begin with a Big Bang), where  $\Omega_{\text{max}} \geq 0$  is an unspecified “hyper-parameter”. Figure 1 shows the dependence of the posterior probability of the accelerating model,  $\text{Pr}(\text{accel.}|d, I)$ , on  $\Omega_{\text{max}}$ . Even the peak values of  $\text{Pr}(\text{accel.}|d, I)$  are considerably lower than the posterior fraction quoted above, and the dependence on the unknown value of  $\Omega_{\text{max}}$  is significant as well.

Rather than introducing an arbitrary new parameter, another option is to adopt the two-step method described in Section 4, using some of the SN data to obtain a partial posterior in  $\Omega_m$  and  $\Omega_\Lambda$  for both the accelerating and matter-only models and then using the remainder to perform model comparison. The results of doing so are shown in Figure 2 for several different partitioning options (and assuming the two models are equally probable a priori). These results again illustrate the standard Bayesian result that the better-fitting accelerating model is not favoured so decisively over the more predictive (*i.e.*, “simpler”) matter-only model, a result that is robust to prior choice.

This two-step approach to model comparison could be applied to a variety of problems in astrophysics and cosmology (*e.g.*, Bailer-Jones 2012, Khanin & Mortlock 2014).

## References

- Bailer-Jones, C. 2012 *Astronomy & Astrophysics*, 546, A89  
Cox, R. T. 1946, *American Journal of Physics*, 14, 1  
Drell, P. S., Lored, T. J. & Wasserman, I. 2000, *The Astrophysical Journal*, 530, 593  
Efsthathiou, G. 2008, *Monthly Notices of the Royal Astronomical Society*, 388, 1314  
Jenkins, C. R. & Peacock, J. A. 2011, *Monthly Notices of the Royal Astronomical Society*, 413, 2895  
Khanin, A., Mortlock, D. J. 2014, *Monthly Notices of the Royal Astronomical Society*, 444, 1591  
Lempers, F. B. 1971, *Posterior Probabilities of Alternative Linear Models*, Rotterdam: University Press  
O’Hagan, A. 1995, *Journal of the Royal Statistical Society B*, 30, 490  
Perlmutter, S., *et al.* 1999, *The Astrophysical Journal*, 517, 565  
Spiegelhalter, D. J. & Smith, A. F. M., 1982, *Journal of the Royal Statistical Society B*, 44, 377  
Riess, A. G., *et al.* 1998, *The Astronomical Journal*, 116, 1009